REMARKS

In view of the following discussion, the Applicants submit that none of the claims now pending in the application is anticipated under the provisions of 35 U.S.C. § 102 or made obvious under the provisions of 35 U.S.C. § 103. Thus, the Applicants believe that all of these claims are now in allowable form.

I. REJECTION OF CLAIMS 1, 5-7, 9, 13-16, 20, 22 AND 26-31 UNDER 35 U.S.C. § 102

The Examiner has rejected claims 1, 5-7, 9, 13-16, 20, 22 and 26-31 under 35 U.S.C. §102(a) as being anticipated by the Thrift et al. patent (US patent 6,188,985, issued on February 13, 2001, hereinafter Thrift). In response, the Applicants have amended independent claims 1, 9, 15, 16, 22, 30 and 31, from which claims 5-7, 13-14, 20 and 26-29 depend, to more clearly recite aspects of the present invention.

Thrift teaches a voice-activated device for controlling a processor-based host system (such as a computer connected to the World Wide Web). In one particular embodiment, Thrift teaches a voice-activated remote control that performs at least some voice recognition processing on an input user command and then outputs recognized speech to the host computer. The host computer then interprets the recognized speech for the purpose of executing the user command at the host computer. In some cases, the host computer may dynamically generate the grammar used by the remote control for speech recognition, based on Web pages and links that are currently displayed on the host computer. However, Thrift does not teach that the grammar used by the remote control for speech recognition is dynamically updated based on a local parameter of the input user command (*e.g.*, speech signal).

The Examiner's attention is directed to the fact that Thrift fails to disclose or suggest the novel invention of updating or adapting a speech recognition process based on a local parameter of the speech signal being processed, as claimed in Applicants' independent claims 1, 9, 15, 16, 22, 30 and 31. Specifically, Applicants' claims 1, 9, 15, 16, 22, 30 and 31, as amended, positively recite:

1. Method for performing speech recognition, said method comprising the steps of:

(a) receiving a speech signal locally from a user via a client device;

(b) performing speech recognition on said speech signal in accordance with an embedded speech recognizer of said client device to produce a recognizable text signal, wherein said embedded speech recognizer employs a language model;

(c) adapting said performance of speech recognition based on at least one local parameter of said speech signal; and

(d) forwarding said recognizable text signal to a remote server.. (Emphasis added)

9. Method for performing speech recognition, said method comprising the steps of:

(a) receiving a recognizable text signal representative of a user speech signal from a client device, wherein said recognizable text is generated using a speech recognizer having a language model on said client device, and wherein said recognizable text is generated in accordance with adapting said performance of speech recognition based on at least one local parameter of said speech signal; and

(b) processing said recognizable text signal in accordance with a task model. (Emphasis added)

15. A distributed system for performing speech recognition, said system comprising:

a client device for receiving a speech signal locally from a user, wherein said client device having an embedded speech recognizer with a language model for performing speech recognition on said speech signal to produce a recognizable text signal, and wherein said embedded speech recognizer further adapts said performance of speech recognition based on at least one local parameter of said speech signal; and

a remote server for receiving said recognizable text signal. (Emphasis added)

16. A client device for performing speech recognition, said client device comprising:

means for receiving a speech signal locally from a user;

means for performing speech recognition on said speech signal to produce a recognizable text signal, wherein said speech recognition means employs a language model;

means for adapting said performance of speech recognition based on at

least one local parameter of said speech signal; and

means for forwarding said recognizable text signal to a remote server. (Emphasis added)

22. A server for performing speech recognition, said server comprising:

means for receiving a recognizable text signal representative of a user speech signal from a client device, wherein said recognizable text is generated using a speech recognizer having a language model on said client device, and wherein said recognizable text is generated in accordance with adapting said performance of speech recognition based on at least one local parameter of said speech signal; and

means for processing said recognizable text signal in accordance with a task model. (Emphasis added)

30. A computer-readable medium having stored thereon a plurality of instructions, the plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps comprising of:

(a) receiving a speech signal locally from a user via a client device;

(b) performing speech recognition on said speech signal in accordance with an embedded speech recognizer of said client device to produce a recognizable text signal, wherein said embedded speech recognizer employs a language model;

(c) adapting said performance of speech recognition based on at least one local parameter of said speech signal; and

(d) forwarding said recognizable text signal to a remote server. (Emphasis added)

31. A computer-readable medium having stored thereon a plurality of instructions, the plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps comprising of:

(a) receiving a recognizable text signal representative of a user speech signal from a client device, wherein said recognizable text is generated using a speech recognizer having a language model on said client device, and wherein said recognizable text is generated in accordance with adapting said performance of speech recognition based on at least one local parameter of said speech signal; and

(b) processing said recognizable text signal in accordance with a task model. (Emphasis added)

Applicants' invention is directed to a method and apparatus for providing a dynamic speech-driven control and remote service access system, for example for use in connection with portable devices such as cell phones, pagers, personal digital assistants and the like. Many such portable devices rely at least in part on speech-driven user interfaces, which do not require a great deal of physical space to incorporate in an associated device. However, the small physical size of most portable devices also typically limits the processing power, and hence the robustness of a speech recognition system, that can be incorporated in a portable device. Thus, the processing demands of speech processing and recognition often exceed the processing capabilities of typical portable devices.

The present invention provides a method and apparatus for improving the speech processing and recognition capabilities of a client device (*e.g.*, a portable device) through interaction with a central server. In one embodiment, the client device is equipped with an initial language model for speech recognition that facilitates recognition of top-level user requests (*e.g.*, general inquiries) and/or local "speaker adaptation" (*e.g.*, adaptation of local parameters such as environmental noise and speaker pronunciation). As the client device interacts with a user, the central server updates the client device's language model as necessary, *e.g.*, by providing more tailored language models (*e.g.*, beyond the initial language model) when required for processing the user's input. This distributed approach maximizes the processing power of the client device without overburdening the client device unnecessarily with a computationally complex language model, thus providing the client device with just enough data and information to perform the tasks required by the user.

In contrast, Thrift only teaches a speech recognition grammar for remote control of a host that is dynamically generated based on a current display (*e.g.*, of Web pages or links) on the host. In other words, the dynamically generated grammar is based on what the user might say (*e.g.*, valid commands based on the current display), and not on what the user has already said in an input speech signal.

The Applicants' invention positively claims the step of adapting the speech recognition process based on at least one local parameter of the input speech signal

(e.g., environmental noise, user pronunciation and the like). This allows the server to update the client device's language model dynamically, as the client device interacts with a user providing spoken input. Thus, the client device's language model is updated by the server as necessary with just enough data and information to perform the tasks required by the user, thereby conserving the client device's limited processing and memory capacity. Thrift's system is completely devoid of any teaching relating to the need or desire to dynamically generate a language model based on a feature or local parameter of the input speech signal.

Therefore, the Applicants submit that, at least for the reasons presented above, independent claims 1, 9, 15, 16, 22, 30 and 31, as amended, fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

Dependent claims 5-7, 13-14, 20 and 26-29 depend from claims 1, 9, 15, 16 and 22 and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 5-7, 13-14, 20 and 26-29 are not anticipated by the teachings of Haskell. Therefore, the Applicants submit that dependent claims 5-7, 13-14, 20 and 26-29 also fully satisfy the requirements of 35 U.S.C. §102 and are patentable thereunder.

II. REJECTION OF CLAIMS 2-4, 10-12, 17-19 AND 23-25 UNDER 35 U.S.C. § 103

The Examiner rejected claims 2-4, 10-12, 17-19 and 23-25 under 35 U.S.C. §103(a) as being unpatentable over Thrift in view the Balakrishnan et al. patent (U.S. Patent No. 6,182,038, issued January 30, 2001, hereinafter Balakrishnan). In response, the Applicants have amended independent claims 1, 9, 15, 16 and 22, from which claims 2-4, 10-12, 17-19 and 23-25 depend, as discussed above to more clearly recite aspects of the invention.

Thrift has been discussed above.

Balakrishnan teaches a computer speech recognition system that operates independent of an application's vocabulary and language models. In particular, Balakrishnan teaches a method for generating context-dependent (CD) phoneme networks as an intermediate speech recognition step. These CD phoneme networks

are generated from acoustic models and are specific to a user and environment. A user's CD phoneme network may then be provided to an application having an independent vocabulary and language model. Thus, final speech recognition is performed at the application in accordance with an application-specific language model. However, Balakrishnan, like Thrift, does not teach that the language model used by the application for speech recognition is dynamically updated based on a local parameter of an input speech signal being processed, as claimed in Applicants' independent claims 1, 9, 15, 16 and 22, which have been recited above.

As discussed above, the Applicants' Invention provides a method and apparatus for improving the speech processing and recognition capabilities of a client device (*e.g.*, a portable device) through interaction with a central server. As the client device interacts with a user, the central server updates the client device's initial language model as necessary, *e.g.*, by providing more tailored language models (*e.g.*, beyond the initial language model) when required for processing the user's input. This distributed approach maximizes the processing power of the client device by providing the client device with just enough data and information to perform the tasks required by the user.

In contrast, neither Thrift nor Balakrishnan teaches dynamically adapting a speech recognition process in accordance with an initial language model based on at least one local parameter of the input speech signal (*e.g.*, environmental noise, user pronunciation and the like), as positively recited in Applicants' claims 1, 9, 15, 16 and 22. As discussed above, this allows the server to update the client device's language model dynamically, as the client device interacts with a user providing spoken input.

Moreover, there is no suggestion or motivation to combine Thrift and Balakrishnan in a manner that would yield the claimed invention. Thrift is directed toward a method of dynamically updating the language model of a remote control device based on changing information displayed on the controlled host device. Balakrishnan teaches a method of generating user- and environment-specific CD phoneme networks from acoustic models for use with an application-specific (*e.g.*, static) language model. Balakrishnan makes it clear that the CD phoneme networks, which can be derived from dynamic acoustic models, are separate and independent

from the language model or models that are actually used for speech recognition processing. Balakrishnan therefore actually teaches away from combination with Thrift, as Thrift teaches dynamically adapting a language model based on current information and Balakrishnan teaches dynamically adapting an acoustic model or phoneme network based on current information for use with a static application language model. Thus, the Applicants respectfully submit that the Examiner is clearly using hindsight to pick and choose elements from the references to support his rejection.

It is impermissible to use the claims as a framework from which to choose among individual references to recreate the claimed invention. *W. L. Gore Associates, Inc. v. Garlock, Inc.*, 220 U.S.P.Q. 303, 312 (1983). Moreover, the mere fact that a prior art structure could be modified to produce the claimed invention would not have made the modification obvious unless the prior art suggested the desirability of the modification. *In re Fritch*, 23 U.S.P.Q. 2d 1780, 1783, Fed. Cir. (1992); *In re Gordon*, 221 U.S.P.Q. 1125, 1127, Fed. Cir. (1984) (emphasis added). The rules applicable for combining references provide that there must be a suggestion from within the references to make the combination. *Uniroyal v. Rudkin-Wiley*, 5 U.S.P.Q. 2d 1434, 1438 (Fed. Cir. 1988); *In re Fine*, 5 U.S.P.Q. 2d at 1599 (emphasis added). Therefore, the teachings of Thrift do not provide any justification for combination with the CD phoneme network methodology of Balakrishnan. Thus, at least for the reasons presented above, independent claims 1, 9, 15, 16 and 22 are not made obvious by the teachings of Thrift in view of Balakrishnan.

Dependent claims 2-4, 10-12, 17-19 and 23-25 depend from claims 1, 9, 15, 16 and 22, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 2-4, 10-12, 17-19 and 23-25 are not made obvious by the teachings of Thrift in view of Balakrishnan. Therefore, the Applicants submit that dependent claims 2-4, 10-12, 17-19 and 23-25 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

III. REJECTION OF CLAIMS 8 AND 21 UNDER 35 U.S.C. § 103

The Examiner rejected claims 8 and 21 under 35 U.S.C. §103(a) as being

unpatentable over Thrift in view the Ramaswamy et al. patent (U.S. Patent No. 6,490,560, issued December 3, 2002, hereinafter Ramaswamy). In response, the Applicants have amended independent claims 1 and 16, from which claims 8 and 21 depend, as discussed above to more clearly recite aspects of the invention.

Thrift has been discussed above.

Ramaswamy teaches a natural language understanding system for verifying the identity of a speaker. In particular, Ramaswamy teaches a method for comparing an input behavior from a speaker speech signal with a behavior model to determine whether the speaker is authorized to interact with the system. Aspects of the behavior that are relevant for the purposes of speaker verification include how a user typically greets the system or what tasks the user typically asks the system to perform. In one embodiment, a user's behavior model may include a language model that is personalized for the particular user and stored as a personal cache. However, Ramaswamy, like Thrift, does not teach that the language model(s) used by the system for speech recognition (*e.g.*, of speaker input) is dynamically updated based on a local parameter of the speaker input (*e.g.*, speech signal).

The Examiner's attention is directed to the fact that Ramaswamy, like Thrift, fails to disclose or suggest the novel invention of updating or adapting a speech recognition process based on a local parameter of the speech signal being processed, as claimed in Applicants' independent claims 1 and 16, which have been recited above.

As discussed above, the Applicants' invention provides a method and apparatus for improving the speech processing and recognition capabilities of a client device (*e.g.*, a portable device) through interaction with a central server. As the client device interacts with a user, the central server updates the client device's initial language model as necessary, *e.g.*, by providing more tailored language models (*e.g.*, beyond the initial language model) when required for processing the user's input. This distributed approach maximizes the processing power of the client device by providing the client device with just enough data and information to perform the tasks required by the user.

In contrast, neither Thrift nor Ramaswamy teaches dynamically adapting a speech recognition process based on at least one local parameter of the input speech

signal (*e.g.*, environmental noise, user pronunciation and the like), as positively recited in Applicants' claims 1 and 16. As discussed above, this allows the server to update the client device's language model dynamically, as the client device interacts with a user providing spoken input.

Moreover, there is no suggestion or motivation to combine Thrift and Ramaswamy in a manner that would yield the claimed invention. Thrift is directed toward a method of dynamically updating the language model of a remote control device based on changing information displayed on the controlled host device. Ramaswamy teaches a method of verifying a speaker identity based on a stored language model tailored to an authorized user's behavior. Ramaswamy therefore actually teaches away from combination with Thrift, as the language model of Ramaswamy is dependent on past data (*e.g.*, stored user behavior patterns); dynamic updates of the language model used by Ramaswamy would defeat the purpose of the invention, because it would provide little or no basis for comparison against the current speaker's behavior (*e.g.*, would provide little behavior of a past authorized speaker/user to match to). Thus, the Applicants respectfully submit that the Examiner is clearly using hindsight to pick and choose elements from the references to support his rejection.

Therefore, the remote control teachings of Thrift do not provide any justification for combination with the speaker verification methodology of Ramaswamy. Thus, at least for the reasons presented above, independent claims 1 and 16 are not made obvious by the teachings of Thrift in view of Ramaswamy.

Dependent claims 8 and 21 depend, respectively, from claims 1 and 16, and recite additional features therefore. As such, and for at least the same reasons set forth above, the Applicants submit that claims 8 and 21 are not made obvious by the teachings of Thrift in view of Ramaswamy. Therefore, the Applicants submit that dependent claims 8 and 21 also fully satisfy the requirements of 35 U.S.C. §103 and are patentable thereunder.

IV.INFORMATION DISCLOSURE STATEMENT

The Examiner has indicated that a copy of PCT Patent Application No. WO

99/08084, cited in the Information Disclosure Statement (IDS) filed December 28, 2001, was not provided with the IDS. The Applicants apologize for this oversight and provide herewith a copy of said PCT application. The Applicants respectfully request that the Examiner provide an updated Form 1449 upon receipt and review of the PCT application.

V. CONCLUSION

Thus, the Applicants submit that all of the presented claims now fully satisfy the requirements of 35 U.S.C. §102 and §103. Consequently, the Applicants believe that all these claims are presently in condition for allowance. Accordingly, both reconsideration of this application and its swift passage to issue are earnestly solicited.

If, however, the Examiner believes that there are any unresolved issues requiring the maintenance of the present final action in any of the claims now pending in the application, it is requested that the Examiner telephone Mr. Kin-Wah Tong, Esq. at (732) 530-9404 so that appropriate arrangements can be made for resolving such issues as expeditiously as possible.

Respectfully submitted,

2/15/05
Date

Kin-Wah Tong, Attorney
Reg. No. 39,400
(732) 530-9404

Moser, Patterson & Sheridan, LLP
595 Shrewsbury Avenue
Shrewsbury, New Jersey 07702